MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Sequential Particle Filter For Multiple Object Tracking

Nam Trung Pham, Karianto Leman, and Teck Wee Chua
Institute for Infocomm Research, Singapore
{ntpham, karianto, tewchua}@i2r.a-star.edu.sg

Abstract

Tracking objects in a scene under different degree of occlusions is still a challenge in computer vision. In this paper, we aim to tackle this problem with a jump observation model that includes both visible objects and occluded objects in a sequential particle filter tracking technique. The technique is extended from the sequential estimation method. An hierarchical structure is introduced to model visible objects and its objects that are occluded. With this structure, we are able to handle the weak measurement observation of the occluded objects and try to track objects under occlusions.

1 Introduction

Occlusion is one of the important and necessary factors to consider in tracking multiple objects in video. This is a challenging task due to the lack of information when objects are occluded. A dynamic template updating mechanism [1] is proposed to adapt with the changes of a single object under occlusions, for example, a person goes behind a tree. In multiple object tracking, the occlusion between objects also needs to be considered. Some methods are proposed to track multiple objects under occlusions by assigning objects into different layers [2, 3]. The idea of these methods is that pixels, segments, or motions that have similar properties are grouped into a layer. In [2], Tao proposed a dynamic background layer model and each moving object are modeled as a foreground layer. In [3], a sequential estimation method is proposed to obtain state estimates at each layer. The occlusion also can be handled by considering many levels of tracking [4]. However, the loss information of objects during occlusions may raise a challenge issue for these approaches.

Some other approaches introduce prior and likelihood functions in multiple object tracking to handle interactions between objects such as occlusions [5, 6, 7]. These techniques do not clearly model objects that are occluded. Thus, they can fail under a large degree of occlusion.

To overcome the occlusion problem, some methods [8, 9] take advantage of using multiple cameras to compensate for the lack of object information when using a single camera . These methods can handle the occlusion efficiently.

In this paper we propose a new approach for occlusion handling based a jump observation model that includes both visible objects and occluded objects in a sequential particle filter. We notice that when person B is occluded by person A, the state of person B can be estimated based on visual observations from person A. It is like a share or jump observation model from B to A. Then, this observation model is embedded in a sequential particle filter. Hence, our method can be used for tracking under half occlusions or total occlusions.

2 Sequential particle filter

Let $X_k = \{X_{k,1}, X_{k,2}, ..., X_{k,N}\}$ be the state of multiple objects, where the state of each object, $X_{k,i} = \{x, y, r_x, r_y\}$, is represented using an ellipse with center $\{x, y\}$, radius $\{r_x, r_y\}$. Multiple object state X_k can be estimated by posterior density $p\left(X_k|Z_{1:k}\right)$ where $Z_{1:k}$ are image frames from time 1 to k. Let denote that the high to low object layers are numerated from 1 to N. Each layer contains one object, and the order of objects in X_k is the same with the order of layers. Then, an heuristic approximation for the posterior density is proposed in [3]

$$p\left(X_k|Z_{1:k}\right) \simeq \prod_{j=1}^{N} p\left(X_{k,j}|Z_{1:k-1}, \bar{Z}_k^j\right) \tag{1}$$

where $\bar{Z}_k^j$ is the image Z_k after removing objects from 1 to $j-1$. With this approximation, we propose here a sequential particle filter to estimate multiple object posterior density $p\left(X_k|Z_{1:k}\right)$ by applying the particle filter [10] sequentially as follows.

With previous multiple object estimation $\hat{X}_{k-1}$, we can determine layers for objects at time k by an assumption that if an object has a larger y-position than other objects, this object will belong to a higher layer than others. This assumption is easy to understand in our scenarios because the direction of the camera is known in advance. Then, a particle filter is applied for the first layer to obtain the state estimation for object 1, $\hat{X}_{k,1}$. After that, the image patch that is occupied by object 1 at $\hat{X}_{k,1}$ is replaced by the background patch. This process is repeated until the state estimation of object N, $\hat{X}_{k,N}$, is obtained. The details of the algorithm is summarized in Figure 1. The likelihood function $p\left(\bar{Z}_k^j|\tilde{X}_{k,j}^{(i)}\right)$ will be defined in Section 3. $q_k\left(\cdot|X_{k-1,j}^{(i)}, \bar{Z}_k^j\right)$ is a sampling function. In this context, we choose $q_k\left(X_{k,j}^{(i)}|X_{k-1,j}^{(i)}, \bar{Z}_k^j\right) = p\left(X_{k,j}^{(i)}|X_{k-1,j}^{(i)}\right)$ where $p\left(X_{k,j}^{(i)}|X_{k-1,j}^{(i)}\right)$ is a state transition density from the dynamic model

$$X_{k,j} = X_{k-1,j} + v_{k,j} \tag{2}$$

where $v_{k,j}$ is a process noise, $v_{k,j} \sim \mathcal{N}\left(0, R_k\right)$.

3 Jump observation model

Let us consider the likelihood function for object j. Assuming the camera is fixed, likelihood function $p\left(\bar{Z}_k^j|X_{k,j}\right)$ is defined based on color measurement model $p\left(\bar{Z}_k^{j,c}|X_{k,j}\right)$ and background measure-

At time $k \geq 1$
Step 1. *Initialization*
From previous multiple object estimation $\hat{X}_{k-1}$, sort objects to layers.
Step 2. *Processing*
$\bar{Z}_k^1$ = current image frame
For each object j (w.r.t the layer order)
Obtain predicted samples from $\left\{X_{k-1,j}^{(i)}\right\}_{i=1}^{L}$
$$\tilde{X}_{k,j}^{(i)} \sim q_k\left(\cdot|X_{k-1,j}^{(i)}, \bar{Z}_k^j\right)$$
Update weights for predicted samples
$$\text{Set } \tilde{w}_{k,j}^{(i)} = \frac{p\left(\bar{Z}_k^j|\tilde{X}_{k,j}^{(i)}\right)p\left(\tilde{X}_{k,j}^{(i)}|X_{k-1,j}^{(i)}\right)}{q_k\left(\tilde{X}_{k,j}^{(i)}|X_{k-1,j}^{(i)}, \bar{Z}_k^j\right)}$$
Normalized weights: $\sum_{i=1}^{L} \tilde{w}_{k,j}^{(i)} = 1$
Resample to get $\left\{w_{k,j}^{(i)}, X_{k,j}^{(i)}\right\}_{i=1}^{N}$
Obtain state estimate $\hat{X}_{k,j}$ from $\left\{w_{k,j}^{(i)}, X_{k,j}^{(i)}\right\}_{i=1}^{N}$
The image patch defined by $\hat{X}_{k,j}$ is replaced by the background image patch to form $\bar{Z}_k^{j+1}$
EndFor

Figure 1. Sequential particle filter for multiple object tracking

ment model $p\left(\bar{Z}_k^{j,b}|X_{k,j}\right)$. Thus, the likelihood function is described as

$$p\left(\bar{Z}_k^j|X_{k,j}\right) = p\left(\bar{Z}_k^{j,b}|X_{k,j}\right)p\left(\bar{Z}_k^{j,c}|X_{k,j}\right) \quad (3)$$

Color likelihood model. State $X_{k,j}$ is divided into $m \times n$ small regions. Each of region l is associated with a small image patch in $\bar{Z}_k^j$. Then, each of these image patches will be determined whether it is occluded by other objects from 1 to $j-1$. This can be done based on state estimations $\hat{X}_{k,1:j-1}$ obtained from the sequential particle filter in Section 2. Likelihood model $p\left(\bar{Z}_k^{j,c}|X_{k,j}\right)$ is defined by the multiplication between likelihood models of small image patches

$$p\left(\bar{Z}_k^{j,c}|X_{k,j}\right) = \prod_{l=1}^{m\times n} p\left(\bar{Z}_k^{j,c,l}|X_{k,j}\right) \quad (4)$$

where $\bar{Z}_k^{j,c,l}$ is the observation for small image patch l. In this paper, we choose $n = m = 2$. Each of image patch l may be occluded by other objects. Thus, $\bar{Z}_k^{j,c,l}$ may be borrowed from other objects. The likelihood function for each image patch is modelled by a jump model as follows

$$p\left(\bar{Z}_k^{j,c,l}|X_{k,j}\right) = \sum_{m=1}^{j} p\left(\bar{Z}_k^{j,c,l,m}|X_{k,j}\right)p\left(h_m|h_j\right) \quad (5)$$

where $\bar{Z}_k^{j,c,l,m}$ is the observation borrowed from object m, $p\left(h_m|h_j\right)$ is the transition probability from object j to object m. Elements in the transition matrix is satisfied $\sum_{m=1}^{j} p\left(h_m|h_j\right) = 1$. Here, we set

$$p\left(h_j|h_j\right) = \alpha \quad (6)$$
$$p\left(h_m|h_j\right) = \frac{(1-\alpha)}{j-1} \quad (7)$$

With this likelihood function, when the image patch is occluded, invisible objects also can use the observation from visible objects. This makes the system can track objects through occlusions.

Let $p_l(u)$ be the color histogram (grayscale color space) of image patch l, and $q_l(u)$ be the color histogram of template of object m at occluding image patch l. The likelihood function $p\left(\bar{Z}_k^{j,c,l,m}|X_{k,j}\right)$ is defined by

$$p\left(\bar{Z}_k^{j,c,l,m}|X_{k,j}\right) = \frac{1}{\sqrt{2\pi\sigma_c^2}} exp\left\{-\frac{D_{j,c,l,m}^2}{2\sigma_c^2}\right\} \quad (8)$$

where σ_c is a variance of noise, $D_{j,c,l,m} = \sqrt{1 - \int \sqrt{p_l(u)q_l(u)}du}$ is the Bhattacharyya distance [11].

Background likelihood model. Similarly to the color likelihood model, the background likelihood model is

$$p\left(\bar{Z}_k^{j,b}|X_{k,j}\right) = \sum_{m=1}^{j} p\left(\bar{Z}_k^{j,b,m}|X_{k,j}\right)p\left(h_m|h_j\right) \quad (9)$$

Let $\bar{Z}_k^{j,b,m}$ be the background subtraction image after borrowing observations from object m. Assuming that pixels s_i in $\bar{Z}_k^{j,b,m}$ are independent

$$\begin{aligned} p\left(\bar{Z}_k^{j,b,m}|X_{k,j}\right) &= \prod_{s_i \in r_k^{j,b,m}} p_f\left(s_i|X_{k,j}\right) \prod_{s_i \notin r_k^{j,b,m}} p_b\left(s_i|X_{k,j}\right) \quad (10) \\ &= \prod_{s_i \in r_k^{j,b,m}} \frac{p_f\left(s_i|X_{k,j}\right)}{p_b\left(s_i|X_{k,j}\right)} \prod_{s_i \in \bar{Z}_k^{j,b,m}} p_b\left(s_i|X_{k,j}\right) \\ &\propto \prod_{s_i \in r_k^{j,b,m}} \frac{p_f\left(s_i|X_{k,j}\right)}{p_b\left(s_i|X_{k,j}\right)} \end{aligned}$$

where $r_k^{j,b,m}$ is the region of image patch defined by $X_{k,j}$. $p_f\left(s_i|X_{k,j}\right)$ is the likelihood that pixel s_i belongs to the foreground, and $p_b\left(s_i|X_{k,j}\right)$ is the likelihood that pixel s_i belongs to background.

$$p_f\left(s_i|X_{k,j}\right) = \frac{1}{\sqrt{2\pi\sigma_b^2}} exp\left\{-\frac{\left(s_i-1\right)^2}{2\sigma_b^2}\right\} \quad (11)$$

$$p_b\left(s_i|X_{k,j}\right) = \frac{1}{\sqrt{2\pi\sigma_b^2}} exp\left\{-\frac{{s_i}^2}{2\sigma_b^2}\right\} \quad (12)$$

From equations $(3, 4, 5, 8, 9, 10)$, we can obtain the likelihood model $p\left(\bar{Z}_k^j|X_{k,j}\right)$.

4 Experimental results

We evaluate the performance of the proposed method in sequences from EPFL [12], CAVIAR [13],

Figure 2. Results of multiple object tracking on EPFL data

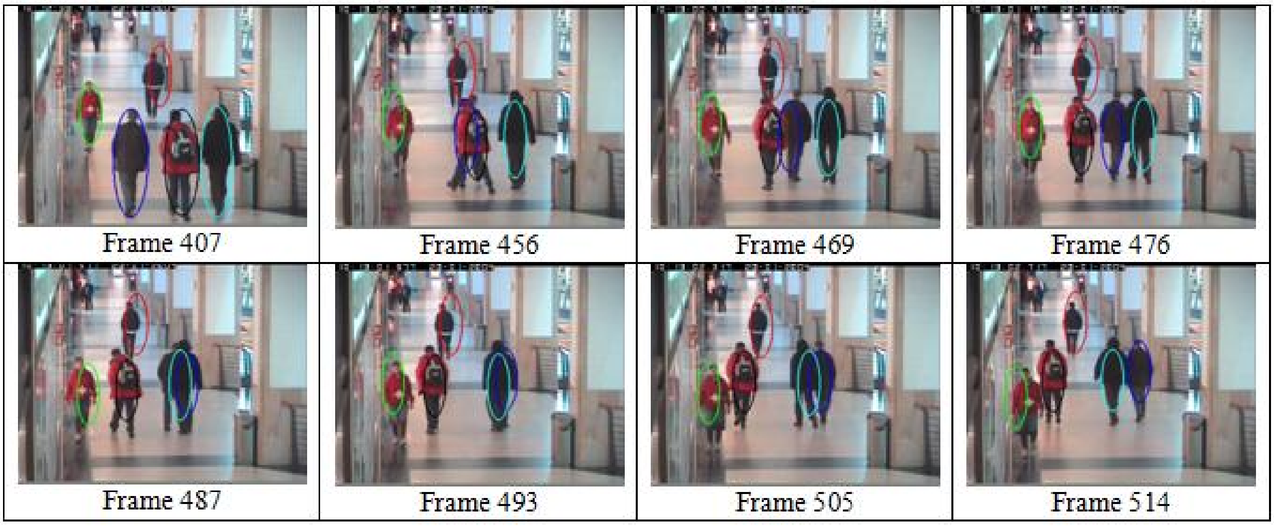

Figure 3. Results of multiple object tracking on CAVIAR data

PETS09 [14]. In these data, our assumption about camera view is satisfied. That means if an object has y-position larger than other ones, this object will belong to the higher layer. We use 100 samples to represent for the multiple object posterior density function. Variance parameters $\sigma_b = 16, \sigma_c = 0.1, \alpha = 0.9$ in the observation model in Section 3 are set by experiments.

In EPFL data, there are about 343 frames. In this scenario, four persons move in the tracking area. Both total and half occlusion appear in this data. Figure 2 shows the comparison between our method and the particle filter with no jump likelihood model. For the particle filter with no jump likelihood model, if persons are far from each other (for examples, frames from 1 to 93), this method can work well. However, when occlusions appear, the tracking performance will be affected (for examples, frame 135). Also in Figure 2, the performance of our method is demonstrated in the first row. In most of time, our method can give good state estimates of objects. This is because of the occluded persons can use the observations from visible persons in our method. Hence, the tracking can maintain well during occlusions.

In CAVIAR data [13], the sequence ThreePastShop2cor.mpg is chosen for evaluating. In this sequence, because the results from background subtraction are not good enough, the detections for new objects are difficult to obtain automatically. Hence, in this case, we assume we have detections for new objects. Then the tracking can be done by the proposed method. In this sequence, the camera is set up at the corridor and look forward. We assume 4 persons are required to track as shown in Figure 3. At frame 456, the black person and red person move to different directions, and they are overlapped together. At frame 487, two black persons are occluded. However, because of the mechanism of our method, observations can be shared between objects. Hence, the proposed method can handle these occlusions. Moreover, when they are near each other, objects are tracked from high to low layers. Hence, the clutter problem that is an important issue for particle filter in visual tracking can be overcome with our framework.

In PETS09 data, we chose S2.L1 sequence for evaluating. This sequence is challenging due to similar appearances and occlusions between persons. The performance of our method in this sequence is shown in Figure 4. We assume that the initial positions of persons are known in advance. Then, our algorithm can give state estimates of these persons. Some challenges

Figure 4. Results of multiple object tracking on PETS data

cases such as three occluding persons from frames 40 to 173, two occluding persons in frames 226, 440, etc. Based on the sequential particle filter with the jump likelihood model, occluded persons can employ observations from visible persons. Hence, tracks of persons are maintained. Figure 4 shows that our method can handle efficiently this scenario.

5 Conclusions

This paper described a method using the jump likelihood model in the sequential particle filter for multiple object tracking. Our sequential particle filter estimates states of objects from high to low layers. The proposed likelihood model can help occluded objects employ observations from visible objects. Hence, our method can provide a mechanism to track objects under occlusions. Moreover, our method is also easy to extend by fusing multiple features. This is based on robust properties of the particle filter in data fusion. Experimental results show that the occlusion can be handled efficiently in our framework.

References

[1] X. Mei and H. Ling, "Robust visual tracking using L1 minimization," in *ICCV*, 2009.

[2] H. Tao, H. S. Sawhney, and R. Kumar, "Dynamic layer representation with applications to tracking," in *CVPR*, 2000.

[3] L. Li, W. Huang, I. Gu, R. Luo, and Q. Tian, "An efficient sequential approach to tracking multiple objects through crowds for real-time intelligent CCTV systems," *IEEE Transaction on Systems, Man, and Cybernetics - Part B*, vol. 38, no. 5, pp. 1254–1268, 2008.

[4] Y. Huang and I. Essa, "Tracking multiple objects through occlusions," in *CVPR*, 2005.

[5] Z. Khan, T. Balch, and F. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets," in *ECCV*, 2003.

[6] K. Smith, D. Perez, and J. M. Odobez, "Using particles to track varying numbers of interacting people," in *CVPR*, 2005, pp. 962 – 969.

[7] T. Zhao and R. Nevatia, "Tracking multiple humans in crowded environment," in *CVPR*, 2004.

[8] S. M. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," in *ECCV*, 2006, pp. 133–146.

[9] K. Kim and L. S. Davis, "Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering," in *ECCV*, 2006, pp. 98–109.

[10] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Transaction on Signal Processing*, 2002.

[11] D. Comaniciu and P. Meer, "Kernel-based object tracking," *PAMI*, vol. 25, no. 5, pp. 564–577, 2003.

[12] "EPFL," http://www.idiap.ch/mucatar.

[13] "CAVIAR," homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/.

[14] "PETS09," http://www.cvg.rdg.ac.uk/PETS2009/a.html.